

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB05/000514

International filing date: 14 February 2005 (14.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: GB

Number: 0403142.3

Filing date: 12 February 2004 (12.02.2004)

Date of receipt at the International Bureau: 16 March 2005 (16.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 7 March 2005



Patent Office
(RECEIVED)**Request for grant of a patent**

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

**The
Patent
Office**



1/77

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

P036769GB

2. Patent application number

(The Patent Office will fill in this part)

12 FEB 2004

0403142.3

3. Full name, address and postcode of the or of each applicant (*underline all surnames*)

ARES TRADING S.A.
Chateau de Vaumarcus
CH-2028 Vaumarcus
Switzerland

13FEB04 EB72920-9 000019
PW/7700 0.00-0403142.3 ACCOUNT CHAPatents ADP number (*if you know it*)

06245682002

If the applicant is a corporate body, give the country/state of its incorporation

Switzerland

4. Title of the invention

PROTEIN

5. Name of your agent (*if you have one*)

Carpmaels & Ransford

"Address for service" in the United Kingdom to which all correspondence should be sent (*including the postcode*)

43-45 Bloomsbury Square
London
WC1A 2RA

Patents ADP number (*if you know it*)

83001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (*if you know it*) the or each application number

Country

Priority application number
(*if you know it*)Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an applicant, or
- c) any named applicant is a corporate body

Yes

See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form.
Do not count copies of the same document

Continuation sheets of this form

Description	49
Claim(s)	7
Abstract	
Drawing(s)	5 + 5

CL

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

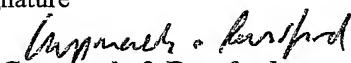
Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination
(*Patents Form 10/77*)

Any other documents
(please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature 
Carpmaels & Ransford Date
12th February 2004

12. Name and daytime telephone number of person to contact in the United Kingdom Hugh R. Goodfellow 020-7242 8692

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.

Protein

This invention relates to a novel protein, termed INSP179, herein identified as a secreted protein, in particular, as a member of the leucine rich repeat containing protein family and to the use of this protein and nucleic acid sequence from the encoding gene in the
5 diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

Background

The process of drug discovery is presently undergoing a fundamental revolution as the era
10 of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

15 As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they
20 become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Introduction

Secreted Proteins

25 The ability for cells to make and secrete extracellular proteins is central to many biological processes. Enzymes, growth factors, extracellular matrix proteins and signalling molecules are all secreted by cells. This is through fusion of a secretory vesicle with the plasma membrane. In most cases, but not all, proteins are directed to the endoplasmic reticulum and into secretory vesicles by a signal peptide. Signal peptides are cis-acting sequences
30 that affect the transport of polypeptide chains from the cytoplasm to a membrane bound

compartment such as a secretory vesicle. Polypeptides that are targeted to the secretory vesicles are either secreted into the extracellular matrix or are retained in the plasma membrane. The polypeptides that are retained in the plasma membrane will have one or more transmembrane domains. Examples of secreted proteins that play a central role in the functioning of a cell are cytokines, hormones, extracellular matrix proteins (adhesion molecules), proteases, and growth and differentiation factors. Description of some of the properties of these proteins follows.

Introduction to Leucine-rich repeat domains

The Leucine-rich repeat (LRR) motif is characterised by tandem arrays of a leucine-rich consensus sequence and a structurally conserved LRR interaction surface that mediate reversible, high affinity protein-protein and protein-RNA interactions (Iozzo, R. V., 1998 *Crit. Rev. Biochem. Mol. Biol.* 32, 141-174; Kajava, A. V. 1998, *J. Mol.Biol.* 277, 519-527; Kobe, B. & Deisenhofer, J. 1995, *Curr. Opin. Struct. Biol.* 5, 409-416). More than 100 LRR-containing proteins have been identified from a diversity of eukaryotic organisms. At least six subfamilies are recognised based upon the differing lengths and consensus sequences of the repeats; 'typical' repeats consist of 20-27 residues. X-ray structural data are currently available for the LRR domains of several proteins including porcine and human placental ribonuclease inhibitor (RI), *Schizosaccharomyces pombe* rna1p, the RNA-binding human spliceosomal U2A' the internalin B protein of the bacterium *Listeria monocytogenes*, and the human mRNA export factor TAP. Although the number of tandem repeats in these LRR structures is variable, ranging from 5 to 17 repeats, the overall non-globular topology of each LRR is strongly similar. Each repeat is a structural unit consisting of a highly conserved β strand packed against a more variable strand, usually helical; these are assembled along a common axis into an arc-shaped structure lined with parallel β strands along the inner surface, an adjoining asparagine or cysteine loop region, and the α -helix, 310 helix, or extended conformation flanking the outer circumference. The solvent exposed parallel β -sheet lining the inner cleft and the curvature of the structure are general features of the LRR motif. Several studies point to regions of the LRR motif as involved in protein and RNA ligand recognition and binding (22, 26, 30-32). However, as each of these studies has targeted only a few of the potential LRR contact points, it is not yet clear whether such limited interactions are universally employed by this diverse protein family to mediate binding (Clark, L.B. *et al.* 2004, *JBC*).

These domains are found in a number of proteins with diverse functions such as immune response, hormone-receptor interactions, enzyme inhibition, vascular repair, cell adhesion and cellular trafficking. Studies have shown the involvement of LRR proteins in early mammalian development, neural development, cell polarization, regulation of gene expression and apoptosis signalling. LRRs may also be critical to the morphology and dynamics of cytoskeleton. Examples of such proteins include CD180 antigen precursor (Lymphocyte antigen 64), MHC class II transactivator (CIITA), Toll-like receptor 6 precursor, Follicle stimulating hormone receptor precursor (FSH-R) (Follitropin receptor), Lutropin-choriogonadotropic hormone receptor precursor, Leucine-rich repeat-containing G protein-coupled receptor 4 precursor, Platelet glycoprotein IX precursor (GPIX) (CD42A), Oligodendrocyte-myelin glycoprotein precursor, Photoreceptor-associated LRR superfamily protein precursor (Retina specific protein PAL) and High affinity nerve growth factor receptor precursor. Interestingly, certain proteins only seem to contain Leucine rich repeat domains such as, Chondroadherin precursor (cartilage formation), Nogo 66 receptor (inhibition of axon growth) and CD14 (immune response).

Thus there is a great need for the identification of novel leucine rich repeat containing proteins, as these proteins are highly likely to be implicated in disease. Identification of such proteins will also be of importance in increasing the understanding of the underlying pathways that lead to the diseases states and associated disease states, mentioned above, and in developing more effective gene and/or drug therapies to treat these disorders.

THE INVENTION

The invention is based on the discovery that the INSP179 polypeptide is a member of the Leucine rich repeat containing protein family.

In one embodiment of the first aspect of the invention, there is provided a polypeptide which:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14;
- (ii) is a fragment thereof which is a member of the leucine rich repeat containing protein family, or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, the polypeptide according to this first aspect of the invention:

(i) comprises the amino acid sequence as recited in SEQ ID NO:10 or SEQ ID NO:14;

(ii) is a fragment thereof which is a member of the leucine rich repeat containing protein family, or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

According to a second embodiment of this first aspect of the invention, there is provided a polypeptide which:

10 (i) consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14;

(ii) is a fragment thereof which is a member of the leucine rich repeat containing protein family, or having an antigenic determinant in common with the polypeptides of (i); or

15 (iii) is a functional equivalent of (i) or (ii).

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the INSP179 exon 1 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as "the INSP179 exon 2 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:6 is referred to hereafter as "the INSP179 exon 20 3 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:8 is referred to hereafter as "the INSP179 exon 4 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:10 is referred to hereafter as "the INSP179 polypeptide".

Although the Applicant does not wish to be bound by this theory, it is postulated that the first 24 amino acids of the INSP179 polypeptide form a signal peptide.

25 The INSP179 exon 1 polypeptide without this postulated signal sequence is recited in SEQ ID NO: 12. The full length INSP179 polypeptide sequence without this postulated signal sequence is recited in SEQ ID NO: 14.

The polypeptide having the sequence recited in SEQ ID NO: 12 is referred to hereafter as "the INSP179 exon 1 mature polypeptide". The polypeptide having the sequence recited in 30 SEQ ID NO: 14 is referred to hereafter as "the INSP179 mature polypeptide".

The term “INSP179 polypeptides” as used herein includes includes the INSP179 exon 1 polypeptide, the INSP179 exon 2 polypeptide, the INSP179 exon 3 polypeptide, the INSP179 exon 4 polypeptide, the INSP179 polypeptide, the INSP179 exon 1 mature polypeptide and the INSP179 mature polypeptide.

5 Preferably, a polypeptide according to any one of the above-described aspects of the invention functions as a member of the leucine rich repeat containing family of proteins.

By “functions as a member of the leucine rich repeat containing family of proteins” we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features within the polypeptides of the leucine rich repeat

10 containing family of proteins.

In a second aspect, the invention provides a purified nucleic acid molecule which encodes a polypeptide of the first aspect of the invention.

Preferably, the purified nucleic acid molecule comprises the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP179 exon 1 polypeptide), SEQ ID NO:3

15 (encoding the INSP179 exon 2 polypeptide), SEQ ID NO:5 (encoding the INSP179 exon 3 polypeptide), SEQ ID NO:7 (encoding the INSP179 exon 4 polypeptide), SEQ ID NO:9 (encoding the INSP179 polypeptide), SEQ ID NO:11 (encoding the INSP179 mature exon 1 polypeptide) and/or SEQ ID NO:13 (encoding the INSP179 mature polypeptide).

The invention further provides that the purified nucleic acid molecule consists of the
20 nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP179 exon 1 polypeptide), SEQ ID NO:3 (encoding the INSP179 exon 2 polypeptide), SEQ ID NO:5 (encoding the INSP179 exon 3 polypeptide), SEQ ID NO:7 (encoding the INSP179 exon 4 polypeptide), SEQ ID NO:9 (encoding the INSP179 polypeptide), SEQ ID NO:11 (encoding the INSP179 mature exon 1 polypeptide) and/or SEQ ID NO:13 (encoding the
25 INSP179 mature polypeptide).

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that
30 contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth

aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to protein members of the leucine rich repeat containing family of proteins of the first aspect of the invention. Preferably, the ligand inhibits the function of a polypeptide of the first aspect of
5 the invention which is a member of the leucine rich repeat containing family of proteins. Ligands to a polypeptide according to the invention may come in various forms, including natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies, structural or
10 functional mimetics of the aforementioned.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or
15 decrease (antagonise) the level of expression of the gene or the activity of the polypeptide.

Importantly, the identification of the function of the INSP179 polypeptides allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease. Ligands and compounds according to the sixth and seventh aspects of the invention may be identified using such methods. These methods are
20 included as aspects of the present invention.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh
25 aspect of the invention, for use in therapy or diagnosis of diseases in which members of the leucine rich repeat containing family of proteins are implicated. Such diseases include cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, fertility disorders, metabolic disorders, AIDS, renal disease, infections and other pathological
30 conditions, particularly those in which leucine rich repeat containing proteins are implicated. These molecules may also be used in the manufacture of a medicament for the treatment such diseases. Leucine rich repeat containing proteins may also be useful in

designing new methods of contraception, as will compounds that bind to these proteins.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

15 A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored in
20 a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a leucine rich repeat containing protein, such as in cell communication and development of cells.

25 In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-
30 acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the

invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or
5 treatment of a disease.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a host cell of the fifth aspect of the invention, or a ligand
10 of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound
15 administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules,
20 ribozymes and ligands, such as antibodies.

The INSP179 polypeptides are members of the leucine rich repeat containing family of proteins and thus have roles in many disease states. Antagonists of the INSP179 polypeptides are of particular interest as they provide a way of modulating these disease states.

25 In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.
30 A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is

also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

5 Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts
10 for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes
15 (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification:
20 Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.* peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to
25 longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is
30 employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For

example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be 5 fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known 10 modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide 15 derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer- 20 RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in 25 polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the 30 amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such

polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

5 The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the INSP179 polypeptides. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid
10 residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New
15 York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified
20 by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides
25 are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the INSP179 polypeptides. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.
30 Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, *i.e.* between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are

substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. Such mutants also include polypeptides in which one or more of the amino acid residues 5 includes a substituent group.

Typically, greater than 30% identity between two polypeptides is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the INSP179 polypeptide, or with active fragments thereof, of greater than 80%. More preferred 10 polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader technology that forms one 15 aspect of the search tools used to generate the Biopendium™ search database may be used (see PCT application WO 01/69507) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the INSP179 polypeptides, are predicted to be members of the leucine rich repeat containing family of proteins, by virtue of sharing significant structural homology with the INSP179 20 polypeptide sequence. By "significant structural homology" is meant that the Inpharmatica Genome Threader predicts two proteins to share structural homology with a certainty of 10% and above.

Polypeptides may be divided into fragments and similarly fragments of functional equivalents may exist. Such fragments are identified by being members of the same protein 25 family as the full-length polypeptide, or having an antigenic determinant in common with the full-length polypeptides.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of a polypeptide or one of the functional equivalents of that polypeptide. The fragments should comprise at least n 30 consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Fragments of full length polypeptides may consist of combinations of 1, 2, 3 or all 4 neighbouring exon sequences in the polypeptide sequences, respectively. For example, such combinations include exons 1 and 2, exons 2 and 3 or exons 1, 2 and 3, and so on. Such fragments are included in the present invention.

- 5 Such fragments may be "free-standing", *i.e.* not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the
- 10 amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or
15 monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

- 20 The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')₂ and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first
25 aspect of the invention.

By "substantially greater affinity" we mean that there is a measurable increase in the affinity for a polypeptide of the invention as compared with the affinity for known secreted proteins.

- 30 Preferably, the affinity is at least 1.5-fold, 2-fold, 5-fold 10-fold, 100-fold, 10³-fold, 10⁴-fold, 10⁵-fold, 10⁶-fold or greater for a polypeptide of the invention than for known secreted proteins such as members of the leucine rich repeat containing family of proteins.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier 5 protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

10 Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, 77-96 in *Monoclonal Antibodies and Cancer Therapy*,

15 Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, *i.e.*, for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies 20 of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu *et al.*, *Proc. Natl. Acad. Sci. USA*, 84, 3439 (1987)), may also be of use.

25 The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones *et al.*, *Nature*, 321, 522 (1986); Verhoeven *et al.*, *Science*, 239, 1534 (1988); Kabat *et al.*, *J. Immunol.*, 147, 1709 (1991); Queen *et al.*, *Proc. Natl. Acad. Sci. USA*, 86, 10029 (1989); Gorman *et al.*, *Proc. Natl. Acad. Sci. USA*, 88, 34181 (1991); and Hodgson *et al.*, *Bio/Technology*, 9, 421 (1991)). The term "humanised antibody", as 30 used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody.

The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is, an antibody having two different antigen binding domains, each domain being directed against a 5 different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554;

10 Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these 15 applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode a polypeptide sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14 and functionally 20 equivalent polypeptides. These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

25 The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic 30 DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for

example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

5 The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as 10 used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript 15 elongation (Nielsen, P.E. *et al.* (1993) Anticancer Drug Des. 8:53-63).

A nucleic acid molecule which encodes a polypeptide of this invention may be identical to the coding sequence of one or more of the nucleic acid molecules disclosed herein.

These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes a polypeptide SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, 20 SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO: 14. Such nucleic acid molecules may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with 25 or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those 30 which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the functional equivalents of the polypeptides of the first aspect of the invention.

Such a nucleic acid molecule may be a naturally-occurring variant such as a naturally-occurring allelic variant, or the molecule may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or
5 organisms.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative
10 or non-conservative amino acid substitutions, deletions or insertions.

The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic
15 oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may
20 be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A
25 fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present
30 invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a

polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

5 The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time
10 of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

15 The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al.* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger
20 (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution
25 comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions
30 used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the

INSP179 polypeptides and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to such coding sequences, or is a nucleic acid molecule that is complementary thereto. In this

5 regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98%, 99% or more identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the INSP179 polypeptides.

10 The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according

15 to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the INSP179 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

20 In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US

25 Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier

30 Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an

equivalent function to that of the INSP179 polypeptide is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 5 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13), are particularly useful probes. Such probes may be labelled with an analytically-detectable reagent to 10 facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such 15 sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended 20 utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman *et al.*, PNAS USA 85, 8998-9002, 1988). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech 25 Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T. *et al.* (1988) 30 Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic., 1, 111-119). Another method which may be used to retrieve unknown sequences is that of

Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

- 5 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of
- 10 sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the
15 present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch
20 Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular
25 genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, *etc.* among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue
30 localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include *in situ* hybridization techniques and nucleotide amplification techniques, such as

PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may 5 be of a temporal, spatial or quantitative nature.

Gene silencing approaches may also be undertaken to down-regulate endogenous expression of a gene encoding a polypeptide of the invention. RNA interference (RNAi) (Elbashir, SM *et al.*, *Nature* 2001, 411, 494-498) is one method of sequence specific post-transcriptional gene silencing that may be employed. Short dsRNA oligonucleotides are 10 synthesised *in vitro* and introduced into a cell. The sequence specific binding of these dsRNA oligonucleotides triggers the degradation of target mRNA, reducing or ablating target protein expression.

Efficacy of the gene silencing approaches assessed above may be assessed through the measurement of polypeptide expression (for example, by Western blotting), and at the 15 RNA level using TaqMan-based methodologies.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

20 The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al.* (*supra*) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, 25 Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, 30 (*supra*). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into

RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

5 10 Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus,

15 15 TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such 20 as Davis *et al.*, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, (*supra*). Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & 25 Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for 30 secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can

be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportI™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, *i.e.*, RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; *i.e.*, to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an

appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication and/or

- 5 endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of
- 10 stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster

- 15 kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA (the

- 20 "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as Drosophila S2 and Spodoptera Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the

- 25 art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, Phytochemistry 30, 3861-3863 (1991).

In particular, all plants from which protoplasts can be isolated and cultured to give whole

- 30 regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and

other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for 5 example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes that can be employed in tk⁻ or aprt[±] cells, respectively.

10 Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and 15 phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed 20 cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

25 Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked 30 immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid

or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med., 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing

5 labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are

10 commercially available, and may be used to synthesise RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp. (Cleveland, OH)).

15 Suitable reporter molecules or labels, which may be used for ease of detection, include radionucleides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further

20 aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be

25 employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as

desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992), Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (1993; DNA Cell Biol. 12:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists

may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Methods for generating detectable signals in the types of assays described herein will be known to those of skill in the art. A particular example is cotransfected a construct expressing a polypeptide according to the invention, or a fragment that is responsible for binding to target, in fusion with the GAL4 DNA binding domain, into a cell together with a reporter plasmid, an example of which is pFR-Luc (Stratagene Europe, Amsterdam, The Netherlands). This particular plasmid contains a synthetic promoter with five tandem repeats of GAL4 binding sites that control the expression of the luciferase gene. When a potential target or ligand is added to the cells, it will bind the GAL4-polypeptide fusion and induce transcription of the luciferase gene. The level of the luciferase expression can be monitored by its activity using a luminescence reader (see, for example, Lehman *et al.* JBC 270, 12953, 1995; Pawar *et al.* JBC, 277, 39243, 2002).

A further preferred method for identifying an agonist or antagonist of a polypeptide of the invention comprises:

- (a) contacting a labelled or unlabeled compound with the polypeptide immobilized on any solid support (for example beads, plates, matrix support, chip) and detection of the compound by measuring the label or the presence of the compound itself; or
- 5 (b) contacting a cell expressing on the surface thereof the polypeptide, by means of artificially anchoring it to the cell membrane, or by constructing a chimeric receptor being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and
- 10 (c) determining whether the compound binds to and activates or inhibits the polypeptide by comparing the level of a signal generated from the interaction of the compound with the polypeptide with the level of a signal in the absence of the compound.

For example, a method such as FRET detection of a ligand bound to the polypeptide in the presence of peptide co-activators (Norris *et al.*, Science 285, 744, 1999) might be used.

In further preferred embodiments, the general methods that are described above may further comprise conducting the identification of agonist or antagonist in the presence of labelled or unlabelled ligand for the polypeptide.

In another embodiment of the method for identifying agonist or antagonist of a polypeptide of the present invention comprises:

determining the inhibition of binding of a ligand to the polypeptide of the invention on any solid or cellular surface thereof, in the presence of a candidate compound under conditions to permit binding to the polypeptide, and determining the amount of ligand bound to the polypeptide. A compound capable of causing reduction of binding of a ligand is considered to be a competitor which may act as an agonist or antagonist. Preferably the ligand is labelled.

More particularly, a method of screening for a polypeptide antagonist or agonist compound comprises the steps of:

- (a) incubating a labelled ligand with a polypeptide according to the invention on any solid support or the cell surface, or a cell membrane containing a polypeptide of the invention.
- 30

- (b) measuring the amount of labelled ligand bound to the polypeptide on the solid support, whole cell or the cell membrane;
- (c) adding a candidate compound to a mixture of labelled ligand and immobilized polypeptide on the solid support, the whole cell or the cell membrane of step (a) and

5 allowing the mixture to attain equilibrium;

- (d) measuring the amount of labelled ligand bound to the immobilized polypeptide or the whole cell or the cell membrane after step (c); and
- (e) comparing the difference in the labelled ligand bound in step (b) and (d), such that the compound which causes the reduction in binding in step (d) is considered to be an agonist

10 or antagonist.

The polypeptides may be found to modulate a variety of physiological and pathological processes in a dose-dependent manner in the above-described assays. Thus, the "functional equivalents" of the polypeptides of the invention include polypeptides that exhibit any of the same modulatory activities in the above-described assays in a dose-dependent manner.

15 Although the degree of dose-dependent activity need not be identical to that of the polypeptides of the invention, preferably the "functional equivalents" will exhibit substantially similar dose-dependence in a given activity assay compared to the polypeptides of the invention.

In certain of the embodiments described above, simple binding assays may be used, in

20 which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this

25 manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using

30 monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the

polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Assay methods that are also included within the terms of the present invention are those that involve the use of the genes and polypeptides of the invention in overexpression or 5 ablation assays. Such assays involve the manipulation of levels of these genes/polypeptides in cells and assessment of the impact of this manipulation event on the physiology of the manipulated cells. For example, such experiments reveal details of signaling and metabolic pathways in which the particular genes/polypeptides are implicated, generate information regarding the identities of polypeptides with which the studied polypeptides interact and 10 provide clues as to methods by which related genes and proteins are regulated.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with 15 the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble 20 receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or 25 bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

30 The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and

enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable 5 pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least 10 about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention.

The term "therapeutically effective amount" as used herein refers to an amount of a 15 therapeutic agent needed to treat, ameliorate, or prevent a targeted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route 20 of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and 25 tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

30 A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the

carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as

5 by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the

10 ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered.

15 Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open

20 sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript

25 from binding to ribosomes. Such oligonucleotides may be administered or may be generated *in situ* from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate

backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

- 5 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion
- 10 of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises

- 15 administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, *i.e.*, an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

- 20 Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, *in vivo* gene therapy does not require isolation and purification of a patient's cells.

- 25 The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For

example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the 5 packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

10 Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

15 Vaccines according to the invention may either be prophylactic (*i.e.* to prevent infection) or therapeutic (*i.e.* to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual

20 receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides

25 are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may

30 include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried

condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Genetic delivery of antibodies that bind to polypeptides according to the invention may 5 also be effected, for example, as described in International patent application WO98/55607.

The technology referred to as jet injection (see, for example, www.powderject.com) may also be useful in the formulation of vaccine compositions.

A number of suitable methods for vaccination and vaccine delivery systems are described 10 in International patent application WO00/29428.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or 15 susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used 20 directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, *Nature*, 324, 163-166 (1986); Bej, *et al.*, *Crit. Rev. Biochem. Molec. Biol.*, 26, 301-334 (1991); Birkenmeyer *et al.*, *J. Virol. Meth.*, 35, 117-126 (1991); Van Brunt, J., *Bio/Technology*, 8, 291-294 (1990)) prior to analysis.

25 In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

30 a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule

of the invention and the probe;

b)contacting a control sample with said probe under the same conditions used in step a);

c)and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from
5 levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

a)obtaining a tissue sample from a patient being tested for disease;

b)isolating a nucleic acid molecule according to the invention from said tissue sample; and

c)diagnosing the patient for disease by detecting the presence of a mutation in the nucleic
10 acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing
15 amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent
20 conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

25 Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product
30 or a single-stranded template molecule generated by a modified PCR. The sequence

determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc. Natl. Acad. Sci. USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by *in situ* analysis (see, for example, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence *in situ* hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250, 559-562 (1990), and Trask *et al.*, Trends, Genet., 7, 149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science (1996), Vol 274, pp 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al.*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 30 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The

substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/25116 (Baldeschweiler *et al.*)
5. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including
10 robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or
15 increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

20 Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above
25 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by
30 combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, such as

by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, 5 ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently 10 or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to 15 distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

- 20 (a) a nucleic acid molecule of the present invention;
- (b) a polypeptide of the present invention; or
- (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule 25 according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic 30 acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease in which

5 members of the leucine rich repeat containing family of proteins are implicated. Such diseases include cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, metabolic disorders, AIDS, renal disease, infections and other pathological conditions. Such kits may also be used for the detection of reproductive disorders including infertility.

10 Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the INSP179 polypeptides.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

15 **Brief description of the Figures**

Figure 1: Top 10 BLASTP hits for INSP179 polypeptide sequence (SEQ ID NO:10) against NCBI-nr

Figure 2: Alignment of second BLASTP hit against INSP179 polypeptide sequence (SEQ ID NO:10)

20 **Figure 3:** CDD domain prediction for INSP179

Figure 4: Signal peptide prediction (SignalP V2.0) for INSP179 polypeptide sequence (SEQ ID NO: 10)

Examples

25 **Example 1: INSP179 Protein BLAST Results**

The INSP179 polypeptide sequence (SEQ ID NO: 10) was used as a protein BLAST query sequence against the NCBI non-redundant sequence database. Figure 1 shows the top results for the BLAST query. The second hit is for a *Mus musculus* hypothetical protein and matches INSP179 with an E value of zero. An alignment of the two proteins is given

30 in Figure 2.

Example 2: Domain prediction

Figure 3 shows the CDD output for the INSP179 polypeptide sequence (SEQ ID NO: 10). The leucine rich repeat in this polypeptide is clearly visible. The INSP179 polypeptide 5 sequence has also been shown by us to have a trans-membrane region (data not shown) using TMHMM (Krogh, A. *et al* 2001, *J Mol Biol* 305(3),567-80).

Example 3: INSP179 signal sequence

10 Figure 4 shows that INSP179 is predicted to possess a signal peptide at the start of the protein. As the SignalP data in Figure 4 clearly shows, the signal peptide cleavage site is thought to be between residues 24 and 25 of the INSP179 polypeptide sequence (Nielsen, H. *et al.* 1997, Protein Engineering, 10, 1-6; Nielsen, H., and Krogh, A.: Prediction of signal peptides and signal anchors by a hidden Markov model. In Proceedings of the Sixth 15 International Conference on Intelligent Systems for Molecular Biology (ISMB 6), AAAI Press, Menlo Park, California, pp. 122-130 (1998)).

List of INSP179 specific sequences:**SEQ ID NO: 1 (INSP179 nucleotide sequence exon 1)**

1 ATGAAAAACC TCTATTCAG AGTCATTACCA ATAGTTATAG GTCTTTATTT
 5 51 TACTGGAATA ATGACAAATG CATCAAGAAA AAGCAATATT TTATTCAATT
 101 CTGAATGCCA ATGGAATGAA TATATTCTGA CAAATTGTTCTTTACCGGA
 151 AAGTGTGATA TACCTGTGGA CATATCACAG ACAGCAGCCA CTGTGGATGT
 201 AAGTTCATAAT TTCTTTAGAG TTCTCTTACA GTCTCACACG AAAAAAGAAG
 251 AGTGGAAAAT AAAACATCTG GACCTCAGTA ACAATCTCAT ATCAAAAATA
 10 301 ACCTTAAGCC CTTTGTCATA TTTACATGCT TTGGAAGTGT TAAACCTCAG
 351 CAACAATGCC ATCCACTCCC TCTCATTGGA TCTACTCAGT CCTAAGTCCT
 401 CATGGGTGAA ACGCCACAGA AGCAGCTCA GAAACAGGTT TCCATTGCTG
 451 AAGGTGCTCA TTCTCAAAG AAATAAAACTC AGTGACACTC CCAAGG

15 SEQ ID NO: 2 (INSP179 polypeptide sequence exon 1)

1 MKNLYFRVIT IVIGLYFTGI MTNASRKSNI LFNSECQWNE YILTNCSFTG
 51 KCDIPVDISQ TAATVDVSFN FFRVLLQSHT KKEEWKIKHL DLSNNLISKI
 101 TLSPFAYLHA LEVLNLSNNA IHSLSLDLLS PKSSWVKRHR SSFRNRFPPLL
 20 151 KVLILQRNKL SDTPKG

SEQ ID NO: 3 (INSP179 nucleotide sequence exon 2)

25 1 GACTGTGGAA ACTGAAGTCA TTGCAGAGTT TGGATCTGTC ATTCAATGGG
 51 ATATTGCAAA TAGGGTGGTC TGATTTTCAC AACTGCCTGC AACTGGAGAA
 101 TCTCTGTTTA AAGAGCAACA AGATATTCAA AATTCCCCCA CAAGCCTTCA
 151 AGGACCTCAA AAAATTACAG

30 SEQ ID NO: 4 (INSP179 polypeptide sequence exon 2)

1 LWKLKSLQSL DLSFNGILQI GWSDFHNCLO LENLCLKSNK IFKIPPPQAFK
 51 DLKKLQ

35 SEQ ID NO: 5 (INSP179 nucleotide sequence exon 3)

1 GTCATAGACC TTAGCAACAA TGCTCTGATT ACCATCCTAC CAATGATGAT
 51 CATAGCTCTA GAATTTCCCC ATCTAGTGGT TGACTTGCT GATAATAACT
 101 GGCAGTGTGA TGATAGTGTG GCAGTCTTTC AAAATTTAT TTCTGAATCC
 151 TGGAGGAAAA AGTGGAAATGT CATTTGCAAC AGGTCTATAG

40 SEQ ID NO: 6 (INSP179 polypeptide sequence exon 3)

1 VIDLSNNALI TILPMMIIAL EFPHLVVDLA DNNWQCDDSV AVFQNFISES
 51 WRKKWNVICN RSIG

45 SEQ ID NO: 7 (INSP179 nucleotide sequence exon 4)

1 GGAGTGAGGA GGCCAACGGG GGCACCTCCCC AGAGCAGGGAT TTCCAGGGAA
 51 51 ACCCGCCTTC CTCCCCATTCA TCTGCATCGC ATGAAAAGCC TCATAAGGAG
 101 CAAAGCAGAG AGGCCCGAGG GAGGAAGGCA CACGGGCATT TCTACTCTGG
 151 GGAAGAAGGC AAAGGCCGGC TCTGGTCTCA GGAAGAAGCA GAGACGGCTG
 201 CCAAGGAGTG TTAGAAGCAC CCGCGATGTG CAGGCTGCCG GCAAAAAAGA
 251 GGACGCTCCC CAGGACCTGG CTCTGGCGGT GTGCCTGTCA GTGTTCATCA
 301 CATTCCCTGT CGCCTTCAGC CTGGGGGCTT TCACAAGGCC TTATGTTGAC
 351 55 AGACTGTGGC AAAAAAAAGTG CCAGAGCAAA AGCCCTGGCC TGGACAACGC
 401 GTATTCAAAC GAGGGCTCT ACGATGACAT GGAAGCTGCG GGGCACACAC
 451 CACACCCAGA GACCCATCTG CGCCAAGTAT TTCCCTCATCT AAGCCTCTAC
 501 GAGAACCCAGA CCCCTTCTG GGTGACACAG CCACACCCAC ACGCCACCGT

551 AATTCCCTGAT AGAAACTCTGG GAAGGAGCAG AAAGGATCCT GGCAGTTCGC
 601 AGAGCCCAGG ACAGTGCAGG GACAACACCG GGGCAGGAAG TGGAAATGAT
 651 GGTGCAGTCT ATTCCATTCT CCAGAGACAT CCACATGCCG GTAACCGTGA
 701 ACTAATGTCA GCAGCGCAGG ACCACATCCA TAGGAATGAT ATTCTCGGAG
 751 AATGGACTTA TGAAACTGTG GCCCAGGAAG AGCCTCTCAG TGCACATTCA
 801 GTGGGCGTCT CTTCTGTAGC TGGCACGTCT CACGCTGTCT CTGGCTCAAG
 851 CCGTTATGAT TCCAATGAAT TAGACCCCTC CCTCTCGGA GAAATAACAG
 901 CTTCCCTCTG TAAAATGCTA ACACATGCAG AAGCACAGAG GACTGGAGAT
 951 AGTAAGGAAA GAGGGGGCAC TGAAACAGTCA CTTTGGGACT CGCAGATGGA
 10 1001 ATTTTCTAAG GAAAGGCAAG TGAGTCATC CATTGATTG CTGAGCATAAC
 1051 AGCAGCCAAG GCTGTCCGGG GCAAGGGCTG AGGAAGCGCT TTCAGCCCAC
 1101 TACAGCGAGG TTCCATACGG TGACCCAAGA GACACAGGCC CATCAGTCTT
 1151 TCCTCCAAGA TGGGACAGTG GCCTGGATGT CACTCCTGCT AACAAAGGAAC
 1201 CAGTGCAGAA ATCCACTCCT TCTGACACTT GCTGTGAGTT GGAGAGTGAC
 15 1251 TGTGACTCTG ATGAGGGGTC TCTGTTCACT CTGAGCTCCA TAAGTTCAGA
 1301 GAGTGCAAGG AGCAAGACTG AAGAGGCAGT GCCTGTGAG GAGTCCCTGC
 1351 AGGACGAGAG CTCAGGGGCA AGCAAGGACA ATGTGACGGC TGTAGACAGT
 1401 CTTGAGGAAA ATGTTACCTT CCAAACAATT CCAGGGAAAT GCAAGAACATCA
 1451 AGAAGATCCC TTTGAAAAAC CTCTCATTTC TGCTCCAGAC TCTGGCATGT
 20 1501 ACAAGACTCA TCTGGAAAAT GCCTCTGACA CTGATAGATC TGAGGGCCTG
 1551 TCACCCTGGC CCAGGTCACC AGGGAATAGT CCCTTAGGGG ATGAGTTCC
 1601 GGGCATGTTCACTTATGATT ATGACACACAGC TCTTCAATCC AAGGCAGCAG
 1651 AATGGCATTG CTCACTTAGA GACTTAGAAT TTTCAAATGT GGACGTTTA
 1701 CAGCAAACAC CACCATGTTC TGCTGAAGTT CCCTCAGATC CTGATAAGGC
 25 1751 TGCC

SEQ ID NO: 8 (INSP179 polypeptide sequence exon 4)

1 SEEANGGTPQ SRISRETRLP PIHLHRMKSL IRSKAERPQG GRHTGISTLG
 30 51 KKAKAGSGLR KKQRRLPRSV RSTRDVQAAG KKEDAPQDLA LAVCLSVFIT
 101 FLVAFSLGAF TRPYVDRWLWQ KKCQSKSPGL DNAYSNEGFY DDMEAAGHTP
 151 HPETHLRQVF PHLSLYENQT PFWVTQPHPH ATVIPDRTILG RSRKDPGSSQ
 201 SPGQCGDNTG AGSGNDGAVY SILQRPHAG NRELMSAAQD HIHRNDILGE
 251 WTYETVAQEE PLSAHSVGVS SVAGTSHAVS GSSRYDSNEL DPSLSEGEITA
 30 301 SLCKMLTHAE AQRTGDSKER GGTEQSQLWDS QMEFSKERQV SSSIDLLSIQ
 35 351 QPRLSGARAE EALSAHYSEV PYGDPRTDGP SFVPPRWDSG LDVTPANKEP
 401 VQKSTPSDTC CELESDCDSD EGSLFTLSSI SSESARSKTE EAVPDEESLQ
 451 DESSGASKDN VTAVDSLEEN VTFQTIPGKC KNQEDPFEKP LISAPDSGMY
 501 KTHLENASDT DRSEGLSPWP RSPGNPLGD EFPGMFTYDY DTALQSAAE
 551 WHCSLRDLEF SNVDVLQQTP PCSAEVPDSP DKAA

SEQ ID NO: 9 (INSP179 nucleotide sequence)

1 ATGAAAAACC TCTATTCAG AGTCATTACC ATAGTTATAG GTCTTTATTT
 45 51 TACTGGAATA ATGACAAATG CATCAAGAAA AAGCAATATT TTATTCAATT
 101 CTGAATGCCA ATGGAATGAA TATATTCTGA CAAATTGTTT CTTTACCGGA
 151 AAGTGTGATA TACCTGTGGA CATATCACAG ACAGCAGCCA CTGTGGATGT
 201 AAGTTTCAAT TTCTTTAGAG TTCTCTTACA GTCTCACACG AAAAAGAAG
 251 AGTGGAAAAT AAAACATCTG GACCTCAGTA ACAATCTCAT ATCAAAAATA
 30 301 ACCTTAAGCC CTTTGCATA TTTACATGCT TTGGAAGTGT TAAACCTCAG
 35 351 CAACAATGCC ATCCACTCCC TCTCATTGGA TCTACTCAGT CCTAAGTCCT
 401 CATGGGTGAA ACGCCACAGA AGCAGCTCA GAAACAGGTT TCCATTGCTG
 451 AAGGTGCTCA TTCTTCAAAG AAATAAAACTC AGTGACACTC CCAAGGGACT
 501 GTGGAAACTG AAGTCATTGC AGAGTTGGA TCTGTCATTC AATGGGATAT
 55 551 TGCAAATAGG GTGGTCTGAT TTTCACAACT GCCTGCAACT GGAGAATCTC
 601 TGTTTAAAGA GCAACAAGAT ATTCAAAATT CCCCCACAAG CCTTCAAGGA
 651 CCTCAAAAAA TTACAGGTCA TAGACCTTAG CAACAAATGCT CTGATTACCA
 701 TCCTACCAAT GATGATCATA GCTCTAGAAT TTCCCCATCT AGTGGTTGAC
 751 TTGGCTGATA ATAACGGCA GTGTGATGAT AGTGTGGCAG TCTTCAAAA
 801 TTTTATTCT GAATCCTGGA GGAAAAAGTG GAATGTCATT TGCAACAGGT

851 CTATAGGGAG TGAGGAGGCC AACGGGGCA CTCCCCAGAG CAGGATTCC
 901 AGGGAAACCC GCCTTCCTCC CATTCATCTG CATCGCATGA AAAGCCTCAT
 951 AAGGAGCAAA GCAGAGAGGC CCCAGGGAGG AAGGCACACG GGCATTCTA
 1001 CTCTGGGAA GAAGGCAAAG GCCGGCTCTG GTCTCAGGAA GAAGCAGAGA
 5 1051 CGGCTGCCAA GGAGTGTAG AAGCACCCGC GATGTGCAGG CTGCCGGCAA
 1101 AAAAGAGGAC GCTCCCCAGG ACCTGGCTCT GGCGGTGTGC CTGTCAGTGT
 1151 TCATCACATT CCTTGTCGCC TTCAGCCTGG GGGCTTCAC AAGGCCTTAT
 1201 GTTGACAGAC TGTGGCAAAA AAAGTGCAG AGCAAAAGCC CTGGCCTGGA
 1251 CAACCGGTAT TCAAACGAGG GCTTCTACGA TGACATGGAA GCTGCGGGGC
 10 1301 ACACACCACA CCCAGAGACC CATCTGCGCC AAGTATTTC TCATCTAACG
 1351 CTCTACGAGA ACCAGACCCC TTTCTGGGTG ACACAGCCAC ACCCACACGC
 1401 CACCGTAATT CCTGATAGAA CTCTGGGAAG GAGCAGAAAG GATCCTGGCA
 1451 GTTCGCGAGAG CCCAGGACAG TGCGGGGACA ACACCGGGGC AGGAAGTGGA
 1501 AATGATGGTG CAGTCTATT CATTCTCCAG AGACATCCAC ATGCCGGTAA
 15 1551 CCGTGAACTA ATGTCAGCAG CGCAGGACCA CATCCATAGG AATGATATTG
 1601 TCGGAGAATG GACTTATGAA ACTGTGGCCC AGGAAGAGCC TCTCAGTGCA
 1651 CATTCACTGG GCGTCTCTTC TGTAGCTGGC ACGTCTCACG CTGTCTCTGG
 1701 CTCAAGCCGT TATGATTCCA ATGAATTAGA CCCTTCCCTC TCCGGAGAAA
 1751 TAACAGCTTC CCTCTGTAAA ATGCTAACAC ATGCAGAACG ACAGAGGACT
 20 1801 GGAGATAGTA AGGAAAGAGG GGGCACTGAA CAGTCACHTT GGGACTCGCA
 1851 GATGGAATT TCTAAGGAAA GGCAAGTGAG TTCATCCATT GATTGCTGA
 1901 GCATACAGCA GCCAAGGCTG TCGGGGGCAA GGGCTGAGGA AGCGCTTCA
 1951 GCCCACTACA GCGAGGTTCC ATACGGTGAC CCAAGAGACA CAGGCCATC
 2001 AGTCTTCCT CCAAGATGGG ACAGTGGCCT GGATGTCACT CCTGCTAAC
 25 2051 AGGAACCAGT GCAGAAATCC ACTCCTCTG ACACCTGCTG TGAGTTGGAG
 2101 AGTGACTGTG ACTCTGATGA GGGGTCTCTG TTCACTCTGA GCTCCATAAG
 2151 TTCAGAGAGT GCAAGGAGCA AGACTGAAGA GGCAGTGCCT GATGAGGAGT
 2201 CCCTGCAGGA CGAGAGCTCA GGGCAAGCA AGGACAATGT GACGGCTGTA
 2251 GACAGTCTTG AGGAAAATGT TACCTTCAA ACAATTCCAG GGAAATGCAA
 30 2301 GAATCAAGAA GATCCCTTG AAAAACCTCT CATTCTGCT CCAGACTCTG
 2351 GCATGTACAA GACTCATCTG GAAAATGCCT CTGACACTGA TAGATCTGAG
 2401 GGCCTGTCAC CCTGGCCAG GTCACCAGGG AATAGTCCCT TAGGGGATGA
 2451 GTTTCCGGGC ATGTTCACTT ATGATTATGA CACAGCTCTT CAATCCAAGG
 2501 CAGCAGAATG GCATTGCTCA CTTAGAGACT TAGAATTTC AAATGTGGAC
 35 2551 GTTTACAGC AAACACCACC ATGTTCTGCT GAAGTCCCT CAGATCCTGA
 2601 TAAGGCTGCC

SEQ ID NO: 10 (INSP179 polypeptide sequence)

40 1 MKNLYFRVIT IVIGLYFTGI MTNASRKSNI LFNSECQWNE YILTNCSFTG
 51 KCDIPVDISQ TAATVDVSFN FFRVLLQSHT KKEEWKIKHL DLSNNLISKI
 101 TLSPFAYLHA LEVNLNSNNA IHSLSLDLLS PKSSWVKRHR SSFRNRFPILL
 151 KVLLQRNKL SDTPKGLWKL KSLQSLDLSF NGILQIGWSD FHNCLQLENL
 201 CLKSNKIFKI PPQAFKDLKK LQVIDLSNNA LITILPMII ALEFPHLVVD
 25 251 LADNNWQCDD SVAVFQNFIS ESWRKWNVI CNRSIGSEEA NGGTPQSRIS
 301 RETRLPPIHL HRMKSLIRSK AERPOGGRHT GISTLGKAK AGSGGLRKKQR
 351 RLPRSVRSTR DVQAAGKKED APQDLALAVC LSVFITFLVA FSLGAFTTRPY
 401 VDRLWQKKCQ SKSPGLDNAY SNEGFYDDME AAGHTPHPET HLRQVFPHLS
 451 LYENQTPFWV TQPHPHATVI PDRTLGRSRK DPGSSQSPGQ CGDNTGAGSG
 50 501 NDGAVYSILQ RPHAGNREL MSAAQDHIIHR NDILGEWTYE TVAQEEPLSA
 551 HSVGVSSVAG TSHAVSGSSR YDSNELDPSL SGEITASLCK MLTHAEAQRT
 601 GDSKERGGTE QSLWDQSMEF SKERQVSSSI DLLSIQQPRL SGARAEELS
 651 AHYSEVPYGD PRDTGPSVFP PRWDGSLDVT PANKEPVQKS TPSDTCELE
 701 SDCDSDEGSIL FTLSSISSES ARSKTEEAVP DEESLQDESS GASKDNVTAV
 75 751 DSLEENVTFQ TIPGKCKNQE DPFEKPLISA PDSGMYKTHL ENASDTDRSE
 801 GLSPWPRSPG NSPLGDEFPG MFTYDYDTAL QSKAAEWHCS LRDLEFSNVD
 851 VLQQTPPCSA EVPSDPDKAA

SEQ ID NO: 11 (INSP179 mature nucleotide sequence exon 1)

1 TCAAGAAAAA GCAATATTTT ATTCAATTCT GAATGCCAAT GGAATGAATA
 51 TATTCTGACA AATTGTTCTT TTACCGGAAA GTGTGATATA CCTGTGGACA
 5 101 TATCACAGAC AGCAGCCACT GTGGATGTAA GTTCAATT CTTTAGAGTT
 151 CTCTTACAGT CTCACACGAA AAAAGAAGAG TGGAAAATAA AACATCTGGA
 201 CCTCAGTAAC AATCTCATAT CAAAAATAAC CTTAAGCCCT TTTGCATATT
 251 TACATGCTT GGAAGTGTAA AACCTCAGCA ACAATGCCAT CCACCTCCCTC
 301 TCATTGGATC TACTCAGTCC TAAGTCCTCA TGGGTGAAAC GCCACAGAAG
 10 351 CAGCTTCAGA AACAGGTTTC CATTGCTGAA GGTGCTCATT CTTCAAAGAA
 401 ATAAACTCAG TGACACTCCC AAGG

SEQ ID NO: 12 (INSP179 mature polypeptide sequence exon 1)

15 1 SRKSNILFNS ECQWNEYILT NCSFTGKCDI PVDISQTAAT VDVSFNFFRV
 51 51 LLQSHTKKEE WKIKHLDLSN NLISKITLSP FAYLHALEV LNSNNAIHSL
 101 101 SLDLLSPKSS WVKRHRSSFR NRFPLLKVLI LQRNKLSDTP KG

SEQ ID NO: 13 (INSP179 mature nucleotide sequence)

20 1 TCAAGAAAAA GCAATATTTT ATTCAATTCT GAATGCCAAT GGAATGAATA
 51 51 TATTCTGACA AATTGTTCTT TTACCGGAAA GTGTGATATA CCTGTGGACA
 101 101 TATCACAGAC AGCAGCCACT GTGGATGTAA GTTCAATT CTTTAGAGTT
 151 151 CTCTTACAGT CTCACACGAA AAAAGAAGAG TGGAAAATAA AACATCTGGA
 201 201 CCTCAGTAAC AATCTCATAT CAAAAATAAC CTTAAGCCCT TTTGCATATT
 251 251 TACATGCTT GGAAGTGTAA AACCTCAGCA ACAATGCCAT CCACCTCCCTC
 301 301 TCATTGGATC TACTCAGTCC TAAGTCCTCA TGGGTGAAAC GCCACAGAAG
 351 351 CAGCTTCAGA AACAGGTTTC CATTGCTGAA GGTGCTCATT CTTCAAAGAA
 401 401 ATAAACTCAG TGACACTCCC AAGGGACTGT GGAAACTGAA GTCATTGCAG
 451 451 AGTTTGGATC TGTCAATTCAA TGGGATATTG CAAATAGGGT GGCTGTGATT
 501 501 TCACAACCTGC CTGCAACTGG AGAACATCTG TTTAAAGAGC AACAAAGATAT
 551 551 TCAAAATTCC CCCACAAGCC TTCAAGGACC TCAAAAAATT ACAGGTCATA
 601 601 GACCTTAGCA ACAATGCTCT GATTACCATC CTACCAATGA TGATCATAGC
 651 651 TCTAGAATTTC CCCCCATCTAG TGGTTGACTT GGCTGATAAT AACTGGCAGT
 35 701 701 GTGATGATAG TGTGGCAGTC TTTCAAAATT TTATTCTGA ATCTGGAGG
 751 751 AAAAAGTGGA ATGTCATTG CAACAGGTCT ATAGGGAGTG AGGAGGCCAA
 801 801 CGGGGGCACT CCCCAGAGCA GGATTTCAG GGAAACCCGC CTTCCCTCCCA
 851 851 TTCATCTGCA TCGCATGAAA AGCCTCATAA GGAGCAAAGC AGAGAGGCC
 901 901 CAGGGAGGAA GGCACACGGG CATTCTACT CTGGGGAGAG AGGCAAAGGC
 951 951 CGGCTCTGGT CTCAGGAAGA AGCAGAGACG GCTGCCAAGG AGTGTAGAA
 1001 1001 GCACCCCGCA TGTGCAGGCT GCGGCCAAAA AAGAGGACGC TCCCCAGGAC
 1051 1051 CTGGCTCTGG CGGTGTGCCT GTCAGTGTTC ATCACATTCC TTGTCGCCTT
 1101 1101 CAGCCTGGGG GCTTTCACAA GGCCTTATGT TGACAGACTG TGGAAAAAAA
 1151 1151 AGTGCCAGAG CAAAAGCCCT GGCCTGGACA ACAGCGTATTG AAACGAGGGC
 45 1201 1201 TTCTACGATG ACATGGAAGC TGCAGGGCAC ACACCACACC CAGAGACCCA
 1251 1251 TCTGCGCCAA GTATTCCTC ATCTAAGCCT CTACGAGAAC CAGACCCCTT
 1301 1301 TCTGGGTGAC ACAGCCACAC CCACACGCCA CCGTAATTCC TGATAGAACT
 1351 1351 CTGGGAAGGA GCAGAAAGGA TCCTGGCAGT TCGCAGAGCC CAGGACAGTG
 1401 1401 CGGGGACAAC ACCGGGGCAG GAAGTGGAAA TGATGGTGCA GTCTATTCCA
 50 1451 1451 TTCTCCAGAG ACATCCACAT GCGGGTAACC GTGAACTAAT GTCAGCAGCG
 1501 1501 CAGGACCCACA TCCATAGGAA TGATATTCTC GGAGAAATGGA CTTATGAAAC
 1551 1551 TGTGGCCCAAG GAAAGGCCCTC TCAGTGCACA TTCAGTGGGC GTCTCTTCTG
 1601 1601 TAGCTGGCAC GTCTCACGCT GTCTCTGGCT CAAGCCGTTA TGATTCCAAT
 1651 1651 GAATTAGACC CTTCCCTCTC CGGAGAAATA ACAGCTTCCC TCTGTAAAAT
 1701 1701 GCTAACACAT GCAGAACGCAG AGAGGACTGG AGATAGTAAG GAAAGAGGGG
 1751 1751 GCACTGAACA GTCACCTTG GACTCGCAGA TGGAAATTTC TAAGGAAAGG
 1801 1801 CAAGTGAGTT CATCCATTGA TTTGCTGAGC ATACAGCAGC CAAGGCTGTC
 1851 1851 CGGGGCAAGG GCTGAGGAAG CGCTTTCAGC CCACTACAGC GAGGTTCCAT
 1901 1901 ACGGTGACCC AAGAGACACA GGCCCACAG TCTTCCCTCC AAGATGGGAC
 60 1951 1951 AGTGGCCTGG ATGTCACTCC TGCTAACAAAG GAACCAGTGC AGAAATCCAC

2001 TCCTTCTGAC ACTTGCTGTG AGTTGGAGAG TGACTGTGAC TCTGATGAGG
 2051 GGTCTCTGTT CACTCTGAGC TCCATAAGTT CAGAGAGTGC AAGGAGCAAG
 2101 ACTGAAGAGG CAGTGCCTGA TGAGGAGTCC CTGCAGGACG AGAGCTCAGG
 2151 GGCAAGCAAG GACAATGTGA CGGCTGTAGA CAGTCTGAG GAAAATGTTA
 5 2201 CCTTCCAAAC AATTCCAGGG AAATGCAAGA ATCAAGAAGA TCCCTTTGAA
 2251 AAACCTCTCA TTTCTGCTCC AGACTCTGGC ATGTACAAGA CTCATCTGGA
 2301 AAATGCCCTCT GACACTGATA GATCTGAGGG CCTGTACCCC TGGCCCAGGT
 2351 CACCAGGGAA TAGTCCCTTA GGGGATGAGT TTCCGGGCAT GTTCACTTAT
 10 2401 GATTATGACA CAGCTCTTCA ATCCAAGGCA GCAGAAATGGC ATTGCTCACT
 2451 TAGAGACTTA GAATTTCAA ATGTGGACGT TTTACAGCAA ACACCACCAT
 2501 GTTCTGCTGA AGTTCCCTCA GATCCTGATA AGGCTGCC

SEQ ID NO: 14 (INSP179 mature polypeptide sequence)

15 1 SRKSNILFNS ECQWNEYILT NCSFTGKCDI PVDISQTAAT VDVSFNFFRV
 51 LLQSHTKKEE WKIKHLDLSN NLISKITLSP FAYLHALEV LNLNNNAIHSI
 101 SLDLLSPKSS WVKRHRSSFR NRFPLLKVLI LQRNKLSDDTP KGLWKLKSLQ
 151 SLDLSFNGIL QIGWSDFHNC LQLENLCLKS NKIFKIPQQA FKDLKKLQVI
 201 DLSNNALITI LPMMIIALEF PHLVVVLADN NWQCDDSVAV FQNFISESWR
 251 KKWNVICNRS IGSEEANGGT PQSRISRETR LPPIHHLRMK SLIRSKAERP
 301 QGGRHTGIST LGKKAKAGSG LRKKQRRRLPR SVRSTRDVQA AGKKEDAPQD
 351 LALAVCLSVF ITFLVAFSLG AFTRPYVDRL WQKKCQSKSP GLDNAYSNEG
 401 FYDDMEAAGH TPHPETHLRQ VFPHLSLYEN QTPFWVTQPH PHATVIPDRT
 451 LGRSRKDPGS SQSPGQCGDN TGAGSGNDGA VYSILQRPH AGNRELMSAA
 501 QDHIIHRNDIL GEWTYETVAQ EEPLSAHSVG VSSVAGTSHA VSGSSRYDSN
 551 ELDPSLSGEI TASLCKMLTH AEAQRTGDSK ERGGTEQSLW DSQMEFSKER
 601 QVSSSIDLLS IQQPRLSGAR AEEALSAHYS EVPYGDPRDT GPSVFPPRW
 651 SGLDVTTPANK EPVQKSTPSD TCCELESDCD SDEGSLFTLS SISSESARSK
 701 TEEAVPDEES LQDESSGASK DNVTAVDSLE ENVTFQTIPG KCKNQEDPFE
 751 KPLISAPDSG MYKTHLENAS DTDRSEGLSP WPRSPGNNSPL GDEFPGMFTY
 801 DYDTALQSKA AEWHCSLRDL EFSNVDVLQQ TPPCSAEVPS DPDKAA

CLAIMS

1. A polypeptide, which polypeptide:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14;
- 5 (ii) is a fragment thereof which functions as a member of the leucine rich repeat containing family of proteins, or having an antigenic determinant in common with the polypeptide of (i); or
- (iii) is a functional equivalent of (i) or (ii).

10 2. A polypeptide according to claim 1 which:

- (i) comprises the amino acid sequence as recited in SEQ ID NO: 10 or SEQ ID NO: 14;
- (ii) is a fragment thereof which functions as a member of the leucine rich repeat containing family of proteins, or having an antigenic determinant in common 15 with the polypeptide of (i); or
- (iii) is a functional equivalent of (i) or (ii).

3. A polypeptide according to claim 1 or 2 which consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14.

20 4. A polypeptide which is a functional equivalent according to part (iii) of any of the above claims, characterised in that it is homologous to the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 and/or SEQ ID NO:14 and is a member of the leucine rich repeat containing family of proteins.

25 5. A polypeptide which is a fragment or a functional equivalent as recited in any one of claims 1 to 4, which has greater than 80% sequence identity with the amino acid sequence recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 or with an active fragment thereof, preferably greater than 85%, 90%, 95%, 98% or 99% sequence identity.

6. A polypeptide which is a functional equivalent as recited in any one of claims 1 to 5, which exhibits significant structural homology with a polypeptide having the amino acid sequence recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.
- 5 7. A polypeptide which is a fragment as recited in any one of claims 1-3 and claim 5 having an antigenic determinant in common with the polypeptide of part (i) of any one of claim 1 to claim 3 which consists of 7 or more amino acid residues from the amino acid sequence recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.
- 10 8. A purified nucleic acid molecule which encodes a polypeptide according to any one of the preceding claims.
9. A purified nucleic acid molecule according to claim 8, which comprises the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and/or SEQ ID NO:13, or is a redundant equivalent or fragment 15 thereof.
10. A purified nucleic acid molecule according to claim 8 or claim 9 which consists of the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 and/or SEQ ID NO:13, or is a redundant equivalent or fragment thereof.
- 20 11. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claims 8 to 10.
12. A vector comprising a nucleic acid molecule as recited in any one of claims 8 to 11.
13. A host cell transformed with a vector according to claim 12.
14. A ligand which binds specifically to a member of the leucine rich repeat containing 25 family of proteins according to any one of claims 1 to 7.
15. A ligand according to claim 14, which is an antibody.
16. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1 to 7.
17. A compound according to claim 16 that binds to a polypeptide according to any one of 30 claims 1 to 7 without inducing any of the biological effects of the polypeptide.

18. A compound according to claim 17, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.
19. A polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to 5 claim 13, a ligand according to claim 14 or claim 15, or a compound according to any one of claims 16 to 18, for use in therapy or diagnosis of disease.
- 10 20. A method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claims 1 to 7, or assessing the activity of a polypeptide according to any one of claims 1 to 7, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.
21. A method according to claim 20 that is carried out *in vitro*.
- 15 22. A method according to claim 20 or claim 21, which comprises the steps of: (a) contacting a ligand according to claim 14 or claim 15 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.
23. A method according to claim 20 or claim 21, comprising the steps of:
 - a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic 20 acid molecule according to any one of claims 8 to 11 and the probe;
 - b) contacting a control sample with said probe under the same conditions used in step a); and
 - c) detecting the presence of hybrid complexes in said samples; wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid 25 complex in the control sample is indicative of disease.
24. A method according to claim 20 or claim 21, comprising:
 - a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 8 to 11 and the primer;
 - 30 b) contacting a control sample with said primer under the same conditions used in step

- a); and
- c)amplifying the sampled nucleic acid; and
- d)detecting the level of amplified nucleic acid from both patient and control samples; wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

25. A method according to claim 20 or claim 21 comprising:

- a)obtaining a tissue sample from a patient being tested for disease;
- b)isolating a nucleic acid molecule according to any one of claims 8 to 11 from said tissue sample; and
- c)diagnosing the patient for disease by detecting the presence of a mutation which is associated with disease in the nucleic acid molecule as an indication of the disease.

26. The method of claim 25, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

27. The method of claim 25 or claim 26, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

28. A method according to any one of claims 20 to 27, wherein said disease includes, but is not limited to, cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, metabolic disorders, AIDS, renal disease, infections and other pathological conditions.

29. A method according to any one of claims 20 to 27, wherein said disease is a disease in which members of the leucine rich repeat containing family of proteins are implicated.

30. Use of a polypeptide according to any one of claims 1 to 7 as a member of the leucine

rich repeat containing family of proteins.

31. A pharmaceutical composition comprising a polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to claim 13, a ligand according to claim 14 or claim 15, or a compound according to any one of claims 16 to 18.
5
32. A vaccine composition comprising a polypeptide according to any one of claims 1 to 7 or a nucleic acid molecule according to any one of claims 8 to 11.
33. A polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to
10 claim 13, a ligand according to claim 14 or claim 15, a compound according to any one of claims 16 to 18, or a pharmaceutical composition according to claim 31, for use in the manufacture of a medicament for the treatment of cell proliferative disorders, autoimmune/inflammatory disorders, such as arthritis, cardiovascular disorders, neurological disorders, developmental disorders, metabolic disorders, AIDS, renal
15 disease, infections and other pathological conditions.
34. A polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to claim 13, a ligand according to claim 14 or claim 15, a compound according to any one of claims 16 to 18, or a pharmaceutical composition according to claim 31, for use in
20 the manufacture of a medicament for the treatment of a disease in which members of the leucine rich repeat containing family of proteins are implicated.
35. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to claim
25 13, a ligand according to claim 14 or claim 15, a compound according to any one of claims 16 to 18, or a pharmaceutical composition according to claim 31.
36. A method according to claim 35, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide,
30 nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

37. A method according to claim 35, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the 5 patient is an antagonist.

38. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1 to 7, or the level of expression of a nucleic acid molecule according to any one of claims 8 to 11 in tissue from said patient, wherein 10 altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.

39. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1 to 7, or a nucleic acid molecule according to any one of claims 8 to 11 with 15 one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.

40. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule 20 according to any one of claims 8 to 11; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

41. The kit of claim 40, further comprising a third container holding an agent for digesting unhybridised RNA.

25 42. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 8 to 11.

43. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1 to 7; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

30 44. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1 to 7.

45. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 44 with a candidate compound and determining the effect of the compound on the disease of the animal.
46. A polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to claim 13, a ligand according to claim 14 or claim 15, a compound according to any one of claims 16 to 18, or a pharmaceutical composition according to claim 31 for use in IVF, the treatment of fertility related disorders or as a contraceptive.
47. A polypeptide according to any one of claims 1 to 7, a nucleic acid molecule according to any one of claims 8 to 11, a vector according to claim 12, a host cell according to claim 13, a ligand according to claim 14 or claim 15, a compound according to any one of claims 16 to 18, or a pharmaceutical composition according to claim 31 for use in the manufacture of a contraceptive.

Figure 1: Top 10 BLASTP hits for INSP179 polypeptide sequence (SEQ ID NO:10) against NCBI-nr

BLASTP 2.2.2 [Jan-08-2002]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= INSP179.pp
(880 letters)

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,594,288 sequences; 522,190,286 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
ref XP_291099.2 similar to hypothetical protein MGC38937 [Homo ...	1747	0.0
ref NP_705796.1 hypothetical protein MGC38937 [Mus musculus] >g...	684	0.0
ref XP_223356.1 similar to hypothetical protein MGC38937 [Rattu...	650	0.0
ref XP_221358.2 similar to Carboxypeptidase N 83 kDa chain (Car...	82	5e-14
ref NP_570843.1 leucine rich repeat containing 15; leucine-rich...	77	2e-12
gb AAH25836.1 1300018K11Rik protein [Mus musculus]	75	7e-12
ref XP_148373.1 RIKEN cDNA 1300018K11 [Mus musculus] >gi 380806...	75	7e-12
ref NP_659551.1 leucine-rich repeat protein induced by beta-amy...	74	1e-11
dbj BAD01045.1 toll-like receptor 3 [Paralichthys olivaceus]	73	3e-11
dbj BAB85498.1 18 wheeler [Bombyx mori]	72	5e-11



Figure 2: Pairwise alignment of INSP179 to top annotated BLASTP hit (mouse ortholog)

```

>ref|NP_705796.1| hypothetical protein MGC38937 [Mus musculus]
gb|AAH31901.1| Hypothetical protein MGC38937 [Mus musculus]
dbj|BAC34797.1| unnamed protein product [Mus musculus]
Length = 872

Score = 684 bits (1765), Expect = 0.0
Identities = 422/888 (47%), Positives = 529/888 (59%), Gaps = 41/888 (4%)

Query: 1 MKNLYFRVITIVIGLYFTGIMTNASRKSNIIFNSECQWNEYILTCNSFTGKCDIPVDISQ 60
M++ Y RV +V GL F +T SRKS++ FN E Q N +L N S +SQ
Sbjct: 1 MRDFYVRVTILVTLGLCFVETVTPSRKSSVSFNPEYQRNGDLLVNWSSIRH-----VSQ 54

Query: 61 TAATVDVSFNFFRVLLQSHTKKEEWKIKHLDLSNNLISKITLSPFAYLHALEVNLNSNA 120
+D SF FFRVL Q HT+KE IK D +++ ISK+TL P A+LHALE+LNLSN A
Sbjct: 55 NTDAMDRSFYFFRVLFQPFHTQER-HIKPPDRTHHRISKVTLDPLAHLHALEILNLNSNA 113

Query: 121 IHSLSLDLLSPKSSWVKRHRSSFRNRFPLLKVLILQRNKLSDTPKGWLKLSLQLSDLASF 180
IH SLD P SS KRH +R P L+VILQRN+LS TPKGLWLKSL+SLDLSF
Sbjct: 114 IHYFSLDQPLPPSSHQKRHGGSHSRLPRLQVILQLQRNQLSGTPKGLWLKSLRSLDLSF 173

Query: 181 NGILQIGWSDFHNCLQLENLCLKSNKIFKIPPQAFKDLKKLQVIDLSNNALITILEPMMII 240
N I+ IG SDFH CLQLE++ LKSNI I P+AFK LKKLQV+DL +NAL T++P++ I
Sbjct: 174 NRIVHIGLSDFHGCLQLESIYLKSNKICTIHPKAKFLKKLQVVDLRSNALTTLVPIVTI 233

Query: 241 ALEFPHLVVVDLADNNWQCDDSVAVFQNFISESWRKWNVICNRSIGSEEANGGTPQSRIS 300
ALE PHL + LADN WQC +S FQN S SWR+ W ICN S+ ++ N T Q R S
Sbjct: 234 ALELPHLELGLADNQWCSESNVNFQNITSSSWREIWKAI CNMSVENKRPNATHQIRKS 293

Query: 301 RETRL---PPIHLHRMKSLIRSKAERPOGGRHTGISTLGKAKAGSG-LRKKQRRLPRSV 356
R+T L PP L KSLI+SKAERPQ G +S LGK+AK G G LR + P +
Sbjct: 294 RDTHLLLSPPSDSL---KSLIQSKAERPQAGMDMHL SALGKEAKDGYGDLRGWMWPQSPVEL 350

Query: 357 RSTRDVQAAGKKEDAPQDLALAVCLSFITFLVAFLGAFTRPYVDRWLWQKKCQSKPGL 416
R ++D Q +K+D P L LA+CLSVFITF+VAF LGAF RPY+DRL Q++C +K PG
Sbjct: 351 RDSQDEQVTDRKDDKPPALELAICLSVFITFVVAFCGLAFARPYIDRLRQQRCNSKRPGS 410

Query: 417 DNAYSNEGFYDDMEAAGHTPHPETHLRQVFPHLSLYENQTFWVTQPHPHATVIPDRTLG 476
DNAYSN+GF+ D+E A H + T L Q HL L ENQ P WV +P PH+ V ++ LG
Sbjct: 411 DNAYSNKGFGHGDIEGAQHMEOGTDLHQTTHHHLSENQNPSWVAEPIPNSAVQSEQMLG 470

Query: 477 RSRKDPGSSQSPGQCGDNTGAGSGND-----GAVYSILQRPHAGNRELMSAAQDHIIHR 530
+ DPG QSP Q D+ + SG+ + L P+A + +S Q H
Sbjct: 471 SNGTDPGHQOSPEQLKDSNESRGDSIVLPSGPVAHLALHGLPNADAHKAISPVQ---HH 527

Query: 531 NDILGEWTYETVAQEEPLSAHSVGSSVAGTSHAVGSSRYDSNELDPSLSGEITASLCK 590
+D L E Y+TVAQE L + SS+ G S ++L PS ++ AS K
Sbjct: 528 HDFLEEAHYDTVAQEYSLIDDVMDRSSITGPLGTFPSSVESRRDDLHPSQPRDVVASFSK 587

Query: 591 MLTHAEAQRTGDSKERGGTEQ-SLWDQSMEFSKERQVSSSIDLLSIQQPRLSGARAEEL 649
L HA + S E G E DSQM S+EROVS+SI L+ QQP G AEE L
Sbjct: 588 TLAHANTREAEGSMETGCPEPLGAMDSQMGSSERQVSNSIRELATQQPSFQGVDAEERL 647

Query: 650 SAHYSEVPYGDPRDTGPSVFPPRWDGSLDVT PANKEPVQKSTPSDTCCELES--DCDSDE 707
S YSEV + DP PS+ PRW SG V PA EPV++ P D +L + + DSDE
Sbjct: 648 SHVYSEVLHNDP---PSL-RPRWGSGHYVIPATGEPVERDAPFDPHYDLVTNYESDSDE 702

Query: 708 GSLFTLSSISSESARSKTEAVPDE-----ESLQDESSGASKDNVTAVDSLEENVTFQTI 762
GSLFTLSS SE RS EE E + L + G KD+VT+ +S+E+ +T Q I
Sbjct: 703 GSLFTLSSSEGSEDTRSLAEQASVENDGTSQPLPSRNLYEYKDSVTSAESVED-LTSQRI 761

Query: 763 PGKCKNQEDPFEKPLISAPDSGMYKTHLENASDTRSEGLSPWPRSPGNPLGDEFPGMF 822
P KC+ QE LIS PDS + +T+ EN S + E S WP+ PG+ E G +
Sbjct: 762 PEKCEAQEAHLRNTLISGPDSCVCETNQENDSSLDPENRSTWPQLPGHKLSHETLGY 821

```

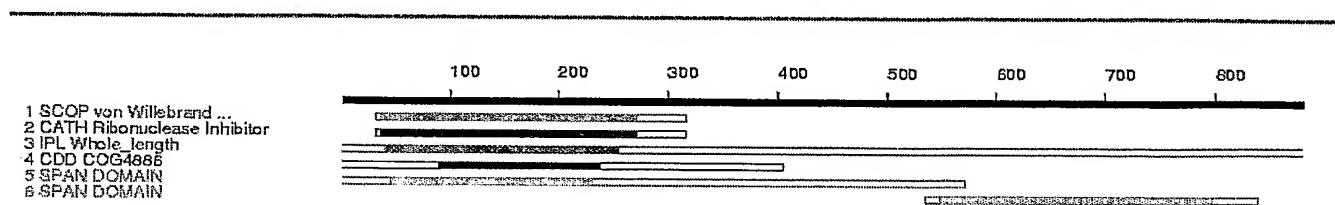


Query: 823 TYDYDTALQSKAAEWHCSLRDLEFSNVDVLQQTTPPCSAEVPSDPDKAA 870
D QS+A +WH SLRDLE NVD +PP S E S P+ A
Sbjct: 822 G---DIEPQSEAVDWHYSLRDLESNVD-SSPSPPYSDEDLSGPEDRA 865



Figure 3: Domain Professor results for INSP179

Query: INSP179.pp



E-values and regions on this page use repsonl and are approximate - hit "AIn" for Blastpgp values and alignments

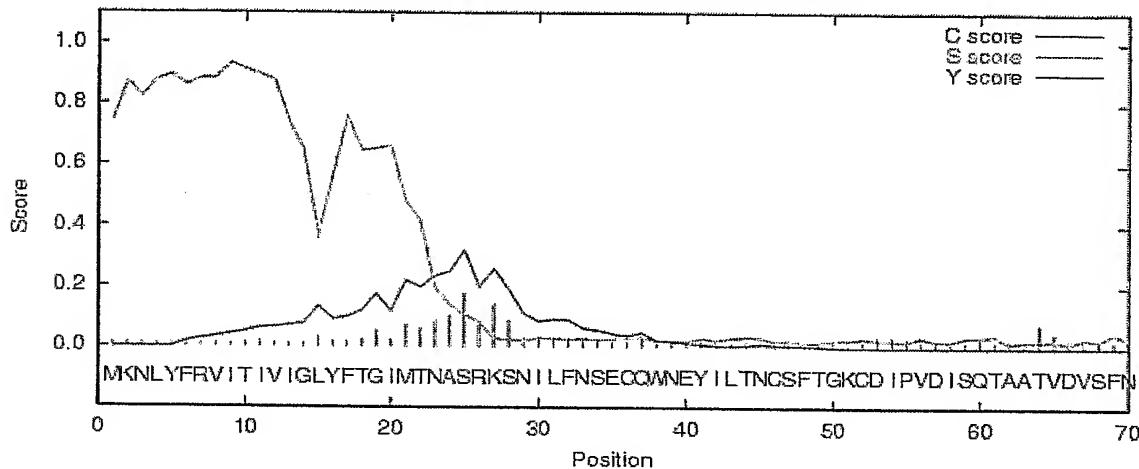
1	SCOP von Willebrand ...	82%	c.10.2.7 (A:) von Willebrand factor binding domain of glycoprotein Ib alpha {Human (Homo sapiens)}	d1gwba	35-271	4-221	- / -	<u>2</u> / 1.4e-25	
2	CATH Ribonuclease Inhibitor	82%	3.80.10.10 Alpha Beta Horseshoe Ribonuclease Inhibitor	1gwba0	35-271	4-221	- / -	<u>2</u> / 1.4e-25	
3	IPL Whole_length	8%	Ferrodoxin-like / Adenylyl and guanylyl cyclase catalytic domain Ascomycotal and Fungal ACs, most of the full length alignment. Have added hCP46367.1 and hCP46942.1 to the alignment for this profile.	IPL002890	40-255	1060-1254	<u>1</u> / 1.4e-15	- / -	
4	CDD COG4886	33%	Leucine-rich repeat (LRR) protein [Function unknown]	COG4886	90-238	98-228	<u>1</u> / 5.3e-14	- / -	
5	SPAN DOMAIN	31%	DOMAIN EXTRACELLULAR (POTENTIAL) sp LGR5_HUMAN O75473 1-561	O75473	45-231	51-222	<u>1</u> / 3.0e-11	- / -	
6	SPAN DOMAIN	81%	DOMAIN ASP/SER-RICH sp SR40_YEAST P32583 25-314	P32583	549-797	15-248	<u>1</u> / 9.0e-05	- / -	



Figure 4: Signal peptide prediction (SignalP V2.0) for INSP179 polypeptide sequence (SEQ ID NO: 10).

>INSP179.pp

SignalP-NN prediction (euk networks): INSP179.pp



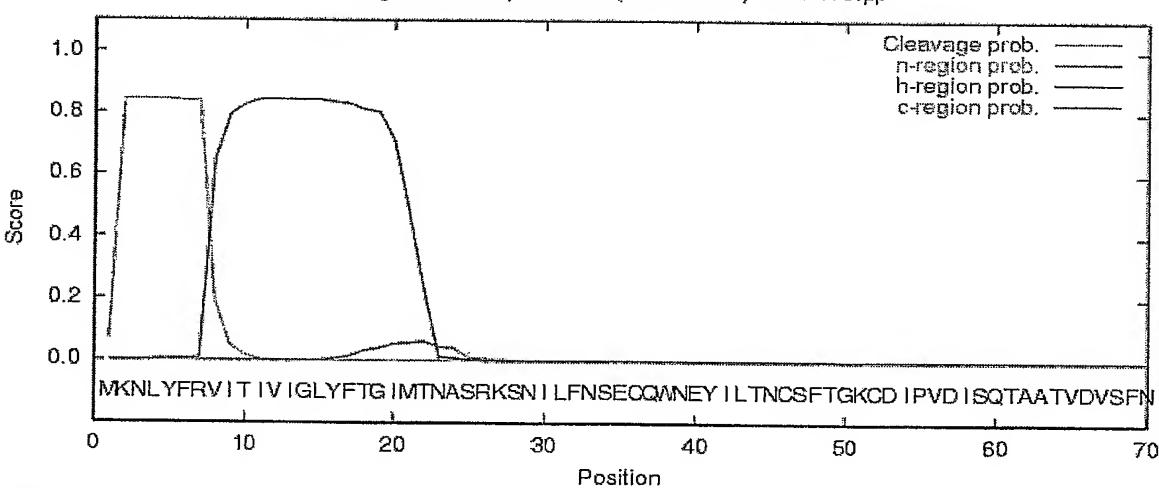
>INSP179.pp

length = 70

# Measure	Position	Value	Cutoff	signal peptide?
max. C	25	0.179	0.33	NO
max. Y	25	0.318	0.32	NO
max. S	9	0.933	0.82	YES
mean S	1-24	0.697	0.47	YES

Most likely cleavage site between pos. 24 and 25: TNA-SR

SignalP-HMM prediction (euk models): INSP179.pp



>INSP179.pp

Prediction: Signal anchor -
Signal peptide probability: 0.065
Signal anchor probability: 0.778

Max cleavage site probability: 0.031 between pos. 24 and 25

